

REMARKS/ARGUMENTS

Applicants have carefully reviewed the above identified application in light of the Office Action dated March 5, 2004. Claims 1-22 remain presented for examination. Claims 1, 7, 13 and 18 have been amended to define still more clearly what Applicants regard as their invention, in terms which distinguish over the art of record.

Claims 1, 7, 14, 16 and 18 are the only independent claims.

Applicants note with appreciation the indication that Claims 3-6 and 9-12 would be allowable if rewritten so as not to depend from a rejected claim, and with no change in scope. These claims have not been so rewritten because, for the reasons given below, their base claim is believed to be allowable.

Claims 1-2, 7-8 and 13-22 were rejected under 35 U.S.C. § 103 as obvious from U.S. Patent 5,812,999 (Tateno).

The present invention, as defined by independent Claim 1, relates to a document descriptor determination method which comprises a step of generalizing input sequences, which input sequences reflecting the structure of the document, to develop general sequences. The method further comprises the step of factoring said input sequences to develop factored sequences. The method then selects a document descriptor from said input sequences, said general sequences, and said factored sequences using minimum descriptor length (MDL) principles. Accordingly, the present invention “relates to inferring (i.e., determining) document descriptors from data within electronic documents” (quoting page 4, lines 9-10 of the specification). Applicants have attempted to make this clearer in the language of the Claim 1 by changing the phrase “document descriptor extraction method” to “document descriptor determination method”.

In this determination process, as described in the specification (inter alia, page 7, line 17 – page 8, line 2), potential descriptors are determined from the generalized sequences and the input sequences. Factoring of the input sequences yields additional potential descriptors. A selection step then selects a descriptor from this group of potential descriptors.

The present invention is illustrated in the specification in terms of inferring document type descriptors (DTDs) from data with eXtensible Markup Language (XML) formatted documents. It is specifically noted that the invention is not so limited, as other language documents are contemplated by the invention (page 4, lines 10-15).

As understood by Applicants, Tateno relates to “an apparatus and method for searching through structured documents and, more particularly, to an apparatus and method for storing the words constituting the text of a structure document provided with tags” (col. 1, lines 8-12). That is, the Tateno reference relates to searching structured documents, using that structure to aide in the search. Tateno relates to analyzing documents for which document descriptors (DTDs) already exist. Thus as indicated in Fig. 1, his process begins with a “text file [3] of a tagged document” (col. 6, lines 30-31). Fig. 2, which represents his searching algorithm, also begins with a reference to processing tagged locations (item 21).

In the Office Action (page 3, first full paragraph), reference is made to Tateno’s Figs. 4 & 5 teaching the input sequences and general sequences of the present invention. However, these figures depict prior art examples of document type definition (DTD) in SGML format and a typical tagged SGML document (some tags having been omitted). Fig. 6 of Tateno, and the accompanying description in the specification, relate to restoring tags that have been removed from the document. But as noted at col. 3, lines 49-63 the resulting document “is acquired by referring to the DTD 40 and thereby restoring the omitted tags” (emphasis mine).

Moreover, as Tateno is working with an electronic document already containing a well-defined structure, it is unclear how the claimed invention’s steps of “factoring” and “minimum descriptor length principles” are applicable to Tateno. Accordingly, Applicants submit that these features of “factoring” and “minimum descriptor length principles” are not obvious over the Tateno reference.

Applicants submit that Tateno fails to teach or suggest the important features of the invention, as defined by claim 1, where document descriptors are determined from input sequences, developed general sequences and developed factored sequences. Accordingly, Applicants submit that Claim 1 is patentable over Tateno. Independent Claims 7 and 18 contain similar features as Claim 1 and are deemed patentable over Tateno for the same reasons.

Independent Claim 16 of the present invention relates to a method for generalizing input sequences to develop general sequences. This method comprises the step of discovering OR patterns among the input sequences. It further comprises the step of discovering sequence patterns among said input sequences and OR patterns. As noted above with the discussion of Claim 1, these features of the invention relate to inferring information related to an electronic document.

In the Office Action rejection of claim 16 (commencing at page 4, last paragraph) Tateno is referenced as containing OR patterns in the DTD 40 document of Fig. 4. Applicants submit, as above, that Tateno's use of a defined DTD document differs from the invention's scope in which such OR patterns are discovered. Moreover, Tateno fails to teach or suggest discovering sequence patterns among input sequences and OR patterns, as these are outside the scope of his analysis of the defined DTD structure.

Consequently, Applicants submit that Tateno fails to teach or suggest the important features of the invention, as defined by claim 16, where general sequences are developed from input sequences. Accordingly, Applicants submit that Claim 16 is patentable over Tateno. Independent Claim 14 contains similar features as Claim 16 and is deemed patentable over Tateno for the same reasons.

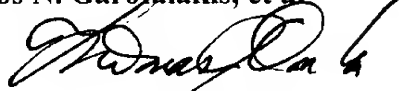
A review of the other art of record has failed to reveal anything which, in Applicants' opinion, would remedy the deficiencies of the art discussed above, as references against the independent claims herein. Those claims are therefore believed patentable over the art of record.

The other claims in this application are each dependent from one or another of the independent claims discussed above and are therefore believed patentable for the same reasons. Since each dependent claim is also deemed to define an additional aspect of the invention, however, the individual reconsideration of the patentability of each on its own merits is respectfully requested.

In view of the foregoing amendments and remarks, Applicants respectfully request favorable reconsideration and early passage to issue of the present application.

Applicants respectfully request that a timely Notice of Allowance be issued in this case.

Respectfully Submitted,
Minos N. Garofalakis, et al



By: Thomas J. Onka, Esq.
Attorney for Lucent

Synnestvedt Lechner & Woodbridge LLP
P.O. Box 592
Princeton, NJ 08542
609-924-3773 phone
609-924-1811 fax